

REGULARIZATION BASED ON ALL-AT-ONCE FORMULATIONS FOR INVERSE PROBLEMS*

BARBARA KALTENBACHER[†]

Abstract. Parameter identification problems typically consist of a model equation, e.g. a (system of) ordinary or partial differential equation(s), and the observation equation. In the conventional reduced setting, the model equation is eliminated via the parameter-to-state map. Alternatively, one might consider both sets of equations (model and observations) as one large system, to which some regularization method is applied. The choice of the formulation (reduced or all-at-once) can make a large difference computationally, depending on which regularization method is used: Whereas almost the same optimality system arises for the reduced and the all-at-once Tikhonov method, the situation is different for iterative methods, especially in the context of nonlinear models. In this paper we will exemplarily provide some convergence results for all-at-once versions of variational, Newton type and gradient based regularization methods. Moreover we will compare the implementation requirements for the respective all-at-one and reduced versions and provide some numerical comparison.

Key words. inverse problems, regularization, all-at-once formulations

AMS subject classifications. 65M32, 65J22, 35R30

1. Introduction. In their original formulation, inverse problems often consist of a model and additional observations. Consider, e.g., an equation (PDE, ODE, integral equation) model for the state u

$$(1) \quad A(x, u) = 0$$

containing a parameter x (or a set of parameters) that is to be determined from additional observations of the state

$$(2) \quad C(u) = y$$

Here $A : \mathcal{D}(A) (\subseteq X \times V) \rightarrow W^*$ and $C : \mathcal{D}(C) (\subseteq V) \rightarrow Y$ are operators acting between Banach spaces X, V, W^*, Y (the star indicates that in variational formulations of models W^* will typically be the dual of some Banach space). The setting could be extended in several directions, e.g., the observation can as well depend on some unknown parameters that have to be identified or the model could consist of a variational inequality instead of an equation. Still (1), (2) is sufficiently general to comprise a wide range of applications, e.g., the following examples.

1.1. Examples.

EXAMPLE 1. Consider a boundary value problem for a linear elliptic PDE on a smooth bounded domain $\Omega \subseteq \mathbb{R}^d$, $d \in \{1, 2, 3\}$

$$-\nabla(a\nabla u) + cu = b \text{ in } \Omega, \quad \frac{\partial u}{\partial n} = g \text{ on } \partial\Omega$$

with a given boundary excitation $g \in H^{-1/2}(\partial\Omega)$, and possibly spatially varying coefficients a, b, c and the inverse problem of identifying these coefficients a, b, c (or part of them) from additional measurements $C(u)$ of the PDE solution u on (part of) the domain or on its boundary. This fits into the above framework with, e.g.,

*This work was supported by the Austrian Science Fund FWF under grant I2271.

[†]Alpen-Adria-Universität Klagenfurt, Austria (barbara.kaltenbacher@aau.at, <http://www.wu.uni-klu.ac.at/bkaltenb/>).

$X = W^{1,p}(\Omega) \times L^\infty(\Omega) \times L^p(\Omega)$, $p \in [1, \infty]$, $V = H^1(\Omega) = W$, (or, if Ω , g are sufficiently smooth, $V = W^{2,p}(\Omega)$, $W^* = L^p(\Omega)$, $p \in [1, \infty]$),

$$\langle A(a, b, c, u), w \rangle_{W^*, W} = \int_{\Omega} (a \nabla u \cdot \nabla w + cuw - bw) dx - \int_{\partial\Omega} gw ds$$

and $Y = L^q(\Sigma)$, $p \in [1, \infty]$, $C(u) = u|_{\Sigma}$, where $\Sigma \subseteq \overline{\Omega}$ is an open subdomain of Ω or a regular curve/surface contained in its boundary or in its interior, so that an embedding or a trace theorem yields continuity of the observation map $C : H^1(\Omega) \rightarrow L^p(\Sigma)$.

EXAMPLE 2. Using similar measurements but a nonlinear model, we consider identification of the nonlinearity, i.e., the function q in the elliptic boundary value problem

$$-\Delta u + q(u) = 0 \text{ in } \Omega, \quad \frac{\partial u}{\partial n} = g \text{ on } \partial\Omega$$

with given g . Here we use a space X that is continuously embedded in $C[\underline{u}, \overline{u}]$ for an interval $[\underline{u}, \overline{u}]$ containing all possibly appearing values of u (which can, e.g., be estimated by using maximum principles in case the PDE is elliptic, depending on the monotonicity of q), $V = H^1(\Omega) = W$,

$$\langle A(q, u), w \rangle_{W^*, W} = \int_{\Omega} (\nabla u \cdot \nabla w + q(u)w) dx - \int_{\partial\Omega} gw ds$$

and Y , C as in Example 1 above.

EXAMPLE 3. Alternatively, one often encounters inverse source problems for nonlinear PDEs such as the simple model example of identifying b in

$$-\Delta u + \zeta u^3 = b \text{ in } \Omega, \quad \frac{\partial u}{\partial n} = g \text{ on } \partial\Omega$$

where g and ζ are given. Here we have $X = L^p(\Omega)$, $p \in [1, \infty]$, $V = H^1(\Omega) = W$,

$$\langle A(b, u), w \rangle_{W^*, W} = \int_{\Omega} (\nabla u \cdot \nabla w + \zeta u^3 w - bw) dx - \int_{\partial\Omega} gw ds,$$

and again Y , C as in Example 1 above.

EXAMPLE 4. Consider identification of the (possibly infinite dimensional) parameter ϑ in the state space system consisting of an ODE model and observations

$$\begin{aligned} \dot{u}(t) &= f(t, u(t), \vartheta) \quad t \in (0, T), \quad u(0) = u_0 \\ y &= C(u), \end{aligned}$$

where the dot denotes time differentiation, $f : (0, T) \times \mathbb{R}^n \times X \rightarrow \mathbb{R}^n$ is a given function and $u_0 \in \mathbb{R}^n$ is a given initial value. Using semigroup theory, this could as well be extended to time dependent PDEs. An example of an infinite dimensional stationary parameter ϑ to be identified in a system of ODEs is the Preisach weight function in some hysteretic evolutionary model. The observations $C(u)$ are, e.g. discrete or continuous in time $y_i = g_i(u(t_i))$, $i \in \{1, \dots, m\}$ (including the case of final measurements $t_i = T$) or $y(t) = g(t, u(t))$, $t \in (0, T)$ with given functions $g_i : \mathbb{R}^n \rightarrow \mathbb{R}^{mM}$ or $g : (0, T) \times \mathbb{R}^n \rightarrow L^2(0, T; \mathbb{R}^M)$.

1.2. Motivation of the all-at-once-approach. Definition and analysis of solution methods for such inverse problems is often based on a reduced formulation that is obtained by the use of a parameter-to-state map i.e., a mapping $S : \mathcal{D}(S) (\subseteq X) \rightarrow V$, that resolves (1) with respect to x

$$\begin{aligned} \forall x \in \mathcal{D}(S) : A(x, S(x)) &= 0 \text{ and} \\ \forall u \in V : ((x, u) \in \mathcal{D}(A) \text{ and } A(x, u) &= 0) \Rightarrow u = S(x). \end{aligned}$$

Existence of such a mapping is guaranteed by the Implicit Function Theorem if for an open set \mathcal{B} with $\mathcal{D}(S) \subset \mathcal{B} \subseteq X$, A is continuously Fréchet differentiable on $\mathcal{B} \times V$ and its derivative A_u with respect to the state is boundedly invertible with uniform bound:

ASSUMPTION 1.

$$\exists C_A \forall (x, u) \in (\mathcal{B} \times V) \cap \mathcal{D}(A) : \|A_u(x, u)^{-1}\| \leq C_A.$$

In order to satisfy this assumption, usually the domain of A has to be restricted, e.g., to

$$(3) \quad \mathcal{D}(A) \subseteq \{(x, u) = (a, b, c, u) : \underline{a} \leq a \leq \bar{a} \text{ a.e. on } \Omega, \ c \geq \underline{c} \text{ a.e. on } \Omega\}$$

with positive constants $0 < \underline{a} < \bar{a}$, $0 < \underline{c}$ in example 1 or to

$$(4) \quad \mathcal{D}(A) \subseteq \left\{ (x, u) = (q, u) : \underline{q} \leq \frac{q(\tilde{\lambda}) - q(\lambda)}{\tilde{\lambda} - \lambda} \ \forall \tilde{\lambda} \neq \lambda \in \mathbb{R} \right\}$$

with some constant $\underline{q} > 0$ in example 2.

Under such conditions the forward operator $F : \mathcal{D}(F) (\subseteq X) \rightarrow Y$, $F = C \circ S$ is well-defined on $\mathcal{D}(F) = \mathcal{D}(S)$ and the inverse problem (1), (2) can be equivalently written as an operator equation

$$(5) \quad F(x) = y$$

Such problems are typically ill-posedness in the sense that F is not continuously invertible and instead of y usually only a noisy version y^δ is available, which we here assume to obey the deterministic and known noise bound δ

$$(6) \quad \|y - y^\delta\| \leq \delta,$$

thus regularization (see, e.g., [2, 6, 15, 17, 20, 21, 24] and the references therein) has to be employed.

In this paper we will return to the original formulation as an all-at-once system of model and observation equation (1), (2),

$$(7) \quad \mathbf{F}(\mathbf{x}) = \mathbf{F}(x, u) = \begin{pmatrix} A(x, u) \\ C(u) \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \mathbf{y}$$

and investigate the behaviour of some well-known regularization paradigms when applied to \mathbf{F} instead of F . We will see that this enables to avoid restrictions like (3), (4) and, moreover, can make a considerable difference when it comes to implementation. We will also provide a convergence analysis that goes beyond the mere application of known results to the operator \mathbf{F} in the sense that regularization might not just be

applied to the whole of $\mathbf{x} = (x, u)$ but - as natural - only to the x part, a case which requires some extra considerations in the convergence analysis.

All-at-once approaches to inverse problems have already been considered previously, see, e.g., [3, 4, 8, 14, 23]. While these papers concentrate on computational aspects and convergence analysis of particular methods, our aim is here to provide a comparative overview on several regularization paradigms.

In the remainder of this paper we will assume that a solution

$$(8) \quad (x^\dagger, u^\dagger) \in \mathcal{D}(\mathbf{F}) \subseteq \mathcal{D}(A) \cap (X \times \mathcal{D}(C))$$

to (7) exists and that $\mathcal{D}(\mathbf{F})$ is convex. Note that $\mathcal{D}(\mathbf{F})$ need not necessarily be the maximal domain of \mathbf{F} , so restriction to a convex set can be done without loss of generality here.

While data misfit and regularization terms will be defined by norms for simplicity of exposition, most of the results are extendable to more general discrepancy and regularization functionals as considered, e.g. in [7, 19, 25].

We treat methods and convergence conditions only exemplarily to highlight analogies and differences between reduced and all-at-once formulations, so our aim is not to provide a complete convergence analysis (a priori and a posteriori choice of regularization parameters, general rates, etc.) for each of the discussed methods.

The remainder of this paper is organized as follows. In Sections 2, 3, 4, we consider Tikhonov, Newton type, and gradient type regularization, respectively. For each of these three paradigms we provide some convergence results with a priori and a posteriori regularization parameter choice strategies and under different assumptions on the forward operator (This analysis part is restricted to just quotation of a convergence result in the gradient method case.) Moreover we compare the key implementation requirements. Section 5 contains a verification of the convergence conditions for the iteratively regularized Gauss-Newton method from Section 3 for Example 3, whose all-at-once and reduced versions are then compared numerically. Some preliminary considerations on comparison of convergence conditions for all-at-once and reduced versions are provided in Section 6, and we give a short summary and an outlook in Section 7.

Notation. For some $r \in [1, \infty)$ we denote by $r^* = \frac{r}{r-1}$ the dual index. $J_r^X = \partial \frac{1}{r} \|\cdot\|_X^r$ denotes the duality mapping with gauge function $\frac{1}{r}t^r$, which is in general set valued. For smooth spaces, i.e., spaces with Gâteaux differentiable norm on the unit sphere, $J_r^X(x)$ will be single valued; otherwise, by a slight abuse of notation, we denote by $J_r^X(x)$ a single valued selection from this set.

2. Tikhonov regularization. For any $\rho, \alpha > 0$, $m, o, r \in [1, \infty)$, $x_0 \in \mathcal{B}$, $u_0 \in V$, we define the pair $(x_\alpha^\delta, u_\alpha^\delta)$ as a minimizer of

$$(9) \quad \min_{(x, u) \in \mathcal{D}(\mathbf{F})} \mathcal{S}(\mathbf{F}(x, u), (0, y^\delta)) + \alpha \mathcal{R}(x, u)$$

with

$$(10) \quad \mathcal{S}((w^*, y), (y^{\text{mod}}, y^{\text{obs}})) = \frac{\rho}{m} \|w^* - y^{\text{mod}}\|_{W^*}^m + \frac{1}{o} \|y - y^{\text{obs}}\|_Y^o$$

and

$$(11) \quad \mathcal{R}(x, u) = \mathcal{R}_1(x) = \frac{1}{r} \|x - x_0\|_X^r$$

or

$$(12) \quad \mathcal{R}(x, u) = \mathcal{R}_2(x, u) = \frac{1}{r} \|x - x_0\|_V^r + \frac{1}{q} \|u - u_0\|_V^q,$$

Note that in case of (11) regularization is only applied to x . In both versions, ρ will remain fixed, whereas α will be chosen in dependence of δ - typically in such a way that it tends to zero as $\delta \rightarrow 0$.

2.1. Well-definedness and convergence. The analysis will be based on the following assumptions

ASSUMPTION 2. *Bounded sets in X , V are weakly (or weakly*) compact.*

ASSUMPTION 3. *\mathbf{F} is weakly (or weakly*) sequentially closed, i.e.,*

$$\begin{aligned} \forall((x_k, u_k))_{k \in \mathbb{N}} \subseteq \mathcal{D}(\mathbf{F}) : \quad & \left(x_k \rightharpoonup x, u_k \rightharpoonup u, A(x_k, u_k) \rightharpoonup f, C(u_k) \rightharpoonup y \right) \\ \implies \quad & \left((x, u) \in \mathcal{D}(\mathbf{F}) \text{ and } A(x, u) = f, C(u) = y \right), \end{aligned}$$

where \rightharpoonup denotes weak or weak* convergence.

For only proving convergence provided minimizers are already well-defined the following somewhat weaker assumption (note that we have strong convergence of the images in the premiss) suffices in place of Assumption 3.

ASSUMPTION 4.

$$\begin{aligned} \forall((x_k, u_k))_{k \in \mathbb{N}} \subseteq \mathcal{D}(\mathbf{F}) : \quad & \left(x_k \rightharpoonup x, u_k \rightharpoonup u, A(x_k, u_k) \rightarrow f, C(u_k) \rightarrow y \right) \\ \implies \quad & \left((x, u) \in \mathcal{D}(\mathbf{F}) \text{ and } A(x, u) = f, C(u) = y \right) \end{aligned}$$

Assumption 2 is satisfied if X, V are reflexive or duals of separable normed spaces. (The star in weak* will be skipped in the following.) Sufficient for Assumption 3 is weak continuity of A and C and weak closedness of $\mathcal{D}(\mathbf{F})$.

To prove well-definedness of a minimizer and convergence in case of regularization with respect to x only, (11), we additionally impose Fréchet differentiability of A and C , Assumption 1 and a growth condition on the derivative A_x of the model operator with respect to the parameter:

ASSUMPTION 5. *There exists a (without loss of generality monotonically increasing) function $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that*

$$(a) \quad \forall(x, v) \in (\mathcal{B} \times V) \cap \mathcal{D}(A) : \|A_x(x, u)\| \leq \psi(\|x\|_X)(1 + \|u\|_V) \\ \text{and } \psi(\|x^\dagger\|_X + \lambda) \lambda \rightarrow 0 \text{ as } \lambda \rightarrow 0 \text{ and } \|x^\dagger - x_0\|_X \text{ is sufficiently small}$$

or

$$(b) \quad \forall(x, v) \in (\mathcal{B} \times V) \cap \mathcal{D}(A) : \|A_x(x, u)\| \leq \psi(\|x\|_X)$$

PROPOSITION 1. *Let Assumptions 2, 3, and in case of (11) additionally Assumptions 1, 5 be satisfied.*

- *Then there exists $\bar{\delta} > 0$ such that for all $\delta \in (0, \bar{\delta})$, $\alpha > 0$ a minimizer of (9) with (10) and (11) or (12) exists.*
- *These minimizers are stable with respect to the data y^δ in the sense that for any sequence $(y_k)_{k \in \mathbb{N}}$ converging to y^δ in the Y norm and any sequence of corresponding minimizers $((x_k, u_k))_{k \in \mathbb{N}}$ there exists a weakly convergent subsequence and the limit of every weakly convergent subsequence is a minimizer of (9).*

- If $\alpha = \alpha(\delta)$ is chosen a priori according to

$$(13) \quad \alpha \rightarrow 0 \text{ and } \frac{\delta^o}{\alpha} \rightarrow 0 \text{ as } \delta \rightarrow 0$$

or a posteriori according to the generalized discrepancy principle

$$(14) \quad \frac{\delta^o}{o} \leq \mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta)) \leq \frac{\tau \delta^o}{o}$$

for some fixed $\tau > 0$ independent of $\delta \in (0, \bar{\delta})$, then there exists $\bar{C} > 0$ independent of $\delta \in (0, \bar{\delta})$ such that for any solution (x^\dagger, u^\dagger) of (1), (2) we have boundedness

$$(15) \quad \forall \delta \in (0, \bar{\delta}) : \frac{1}{r} \|x_{\alpha(\delta)}^\delta - x_0\|_X^r \leq \begin{cases} \frac{1}{r} \|x^\dagger - x_0\|_X^r + \frac{\delta^o}{o\alpha(\delta)} & \text{with (13)} \\ \frac{1}{r} \|x^\dagger - x_0\|_X^r & \text{with (14)} \end{cases}$$

and $\|u_{\alpha(\delta)}^\delta\|_V \leq \bar{C}$

in case of minimizers of (9) with (10) and (11) and

$$(16) \quad \forall \delta \in (0, \bar{\delta}) : \mathcal{R}_2(x_{\alpha(\delta)}^\delta, u_{\alpha(\delta)}^\delta) \leq \begin{cases} \mathcal{R}_2(x^\dagger, u^\dagger) + \frac{\delta^o}{\alpha(\delta)} & \text{with (13)} \\ \mathcal{R}_2(x^\dagger, u^\dagger) & \text{with (14)} \end{cases}$$

in case of minimizers of (9) with (10) and (12).

- In both cases (15), (16), the regularized approximations $(x_{\alpha(\delta)}^\delta, u_{\alpha(\delta)}^\delta)$ converge weakly subsequentially to a solution of (1), (2).

Proof. The assertion follows from the results in Section 3 of [10] with $u, F(u)$ there defined by $(x, u), \mathbf{F}(x, u)$ here. Indeed almost all items of [10, Assumption 2.1] easily follow from Assumption 2.3 (note that it actually suffices to assume weak sequential closedness of the operator in place of [10, Assumption 2.1 (3), (5)]). The only exception arises in case of regularization with respect to x only (11), where boundedness, i.e., weak compactness, of level sets

$$\mathcal{M}_\alpha(M) = \left\{ (x, u) \in \mathcal{D}(\mathbf{F}) : \frac{1}{\alpha} \mathcal{S}(\mathbf{F}(x, u), (0, y^\delta)) + \mathcal{R}(x, u) \leq M \right\}$$

[10, Assumption 2.1 (6)], which the proofs there actually only require to hold for sufficiently small M , has to be shown separately. For obtaining this boundedness, we will make use of Assumptions 1, 5 in that case. Since we have also stated convergence with the discrepancy principle, which is not treated in [10] and for completeness of exposition we provide the details of the convergence part of the proof for the case of (11).

The standard argument of minimality together with (6) for any solution (x^\dagger, u^\dagger) of (1), (2) yields the estimate

$$(17) \quad \mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta)) + \frac{\alpha}{r} \|x_\alpha^\delta - x_0\|^r \leq \frac{1}{o} \delta^o + \frac{\alpha}{r} \|x^\dagger - x_0\|^r$$

thus, upon division by α and setting $\alpha = \alpha(\delta)$ (in case of (14) using the lower estimate there) we get the estimate on x_α^δ in (15). To obtain boundedness of u_α^δ as well, we use the identity

$$(18) \quad A(x_\alpha^\delta, u_\alpha^\delta) = \underbrace{A(x^\dagger, u^\dagger)}_{=0} + \int_0^1 \left(A_x(x^\theta, u^\theta)(x_\alpha^\delta - x^\dagger) + A_u(x^\theta, u^\theta)(u_\alpha^\delta - u^\dagger) \right) d\theta$$

where $x^\theta = x^\dagger + \theta(x_\alpha^\delta - x^\dagger)$, $u^\theta = u^\dagger + \theta(u_\alpha^\delta - u^\dagger)$ and Assumptions 1, 5, as well as (17) or the upper bound in (14)

$$\|A(x_\alpha^\delta, u_\alpha^\delta)\|_{W^*}^m \leq \overline{C}_\delta = \frac{m}{\rho} \begin{cases} \frac{1}{\rho} \delta^\rho + \frac{\alpha(\delta)}{r} \|x_0 - x^\dagger\|_X^r & \text{in case of (13)} \\ \frac{\tau}{\rho} \delta^\rho & \text{in case of (14)} \end{cases}.$$

to arrive at the estimate

$$(19) \quad \begin{aligned} \|u_\alpha^\delta - u^\dagger\|_V &\leq C_A \left(\overline{C}_\delta^{1/m} + \psi(\|x^\dagger\|_X + \|x_\alpha^\delta - x^\dagger\|_X) \|x_\alpha^\delta - x^\dagger\|_X \right) \\ &\times \begin{cases} (1 + \|u^\dagger\|_V + \|u_\alpha^\delta - u^\dagger\|_V) & \text{in case (a)} \\ 1 & \text{in case (b)} \end{cases}. \end{aligned}$$

By the first (already proven) part of (15), this directly yields a bound on $\|u_{\alpha(\delta)}^\delta - u^\dagger\|_V$ in case (b). In case (a) we additionally use the fact that the smallness assumption on $\|x_0 - x^\dagger\|_X$ and the growth condition on ψ allows us to achieve

$$C_A \left(\overline{C}_\delta^{1/m} + \psi(\|x^\dagger\|_X + \varphi(\|x_0 - x^\dagger\|_X)) \varphi(\|x_0 - x^\dagger\|_X) \right) \leq c < 1$$

for some constant c independent of δ , where

$$\varphi(s) = \begin{cases} s + (s^r + \frac{r\delta^\rho}{\rho\alpha(\delta)})^{1/r} & \text{in case of (13)} \\ 2s & \text{in case of (14)} \end{cases}.$$

Rearranging terms in (19) (case (a)) we therefore get

$$\|u_{\alpha(\delta)}^\delta - u^\dagger\|_V \leq \frac{c}{1-c} (1 + \|u^\dagger\|_V).$$

The rest follows by standard arguments from the assumed continuity assumptions on A and C . \square

REMARK 1. Note that in case of regularization with respect to both x and u , (12), we do not need Assumptions 1, 5 and can therefore also deal with situations in which a parameter-to state map not necessarily exists.

Well-definedness of α according to the discrepancy principle (14) follows from [16, Lemma 1] provided $X \times V$ is reflexive and strictly convex and either (i) \mathbf{F} is weakly closed (i.e., Assumption 3 is satisfied) and Y is reflexive or (ii) \mathbf{F} is weak-to-weak continuous and $\mathcal{D}(\mathbf{F})$ is weakly closed (which by the assumed convexity of this set is satisfied, e.g., if it is closed wrt. norm convergence).

If $X \times V$ satisfies the Kadets-Klee property, the results of Theorem 1 imply strong convergence and if the solution is unique then by a subsequence-subsequence argument the whole sequence converges.

2.2. Convergence rates. As usual we can conclude convergence rates with respect to the Bregman distance under source conditions. Just exemplarily we will state a rates result for the all-at-once Tikhonov method with regularization with respect to both x and u under a benchmark source condition. To this end, we use an element of the subgradient

$$(20) \quad \xi \in \partial \mathcal{R}(x^\dagger, u^\dagger) \text{ where } \mathcal{R}(x, u) = \mathcal{R}_2(x, u) = \frac{1}{r} \|x - x_0\|_X^r + \frac{1}{q} \|u - u_0\|_V^q$$

to define the Bregman distance

$$(21) \quad D_{\xi}^{(x_0, u_0)}((x, u), (x^{\dagger}, u^{\dagger})) = \mathcal{R}(x, u) - \mathcal{R}(x^{\dagger}, u^{\dagger}) - \langle \xi, (x - x^{\dagger}, u - u^{\dagger}) \rangle$$

and impose a local smoothness condition on \mathbf{F} .

ASSUMPTION 6.

$$\begin{aligned} & \|A_x(x^{\dagger}, u^{\dagger})(x - x^{\dagger}) + A_u(x^{\dagger}, u^{\dagger})(u - u^{\dagger})\| + \|C'(u^{\dagger})(u - u^{\dagger})\| \\ & \leq \overline{C}_L \left(\frac{\rho}{m} \|A(x, u) - A(x^{\dagger}, u^{\dagger})\|^m + \frac{1}{o} \|C(u) - C(u^{\dagger})\|^o \right)^{1/t} \\ & \quad + LD_{\xi}^{(x_0, u_0)}((x, u), (x^{\dagger}, u^{\dagger})), \end{aligned}$$

for some ξ according to (20), $t > 0$, and all $(x, u) \in \mathcal{D}(\mathbf{F})$.

In case $m = o = t$, Assumption 6 follows from the inverse triangle inequality and the Taylor remainder estimate

$$\begin{aligned} & \|A(x^{\dagger}, u^{\dagger}) + A_x(x^{\dagger}, u^{\dagger})(x - x^{\dagger}) + A_u(x^{\dagger}, u^{\dagger})(u - u^{\dagger}) - A(x, u)\| \\ & + \|C(u^{\dagger}) + C'(u^{\dagger})(u - u^{\dagger}) - C(u)\| \leq LD_{\xi}^{(x_0, u_0)}((x, u), (x^{\dagger}, u^{\dagger})), \end{aligned}$$

which in the quadratic Hilbert space case $D_{\xi}^{(x_0, u_0)}((x, u), (x^{\dagger}, u^{\dagger})) = \frac{1}{2} \|x - x^{\dagger}\|^2 + \frac{1}{2} \|u - u^{\dagger}\|^2$ corresponds to the usual estimate obtained under a local Lipschitz condition on \mathbf{F}' .

PROPOSITION 2. Let \mathbf{F}' satisfy Assumption 6 and let, for some $(v^{\text{mod}}, v^{\text{obs}}) \in W \times Y^*$ with $L \|(v^{\text{mod}}, v^{\text{obs}})\| := L(\|v^{\text{mod}}\|_W + \|v^{\text{obs}}\|_{V^*}) < 1$ the source condition

$$(22) \quad \xi = \mathbf{F}'(x^{\dagger}, u^{\dagger})^*(v^{\text{mod}}, v^{\text{obs}})$$

hold for some $\xi \in \partial \mathcal{R}_2(x^{\dagger}, u^{\dagger})$.

Then for the Tikhonov minimizers according to (9), (10), (12) with the apriori choice

$$(23) \quad \alpha(\delta) \sim \delta^{\frac{o}{t^*}}$$

(with $t^* = \frac{t}{t-1}$) or the a posteriori choice (14) we have

$$(24) \quad D_{\xi}^{(x_0, u_0)}((x_{\alpha(\delta)}^{\delta}, u_{\alpha(\delta)}^{\delta}), (x^{\dagger}, u^{\dagger})) = O(\delta^{o/t})$$

Proof. The proof follows the lines of the classical rates proof from [5].

By (16) and (22) we have in case of (14)

$$\begin{aligned} & D_{\xi}^{(x_0, u_0)}((x_{\alpha(\delta)}^{\delta}, u_{\alpha(\delta)}^{\delta}), (x^{\dagger}, u^{\dagger})) \\ & \leq -\langle (v^{\text{mod}}, v^{\text{obs}}), \mathbf{F}'(x^{\dagger}, u^{\dagger})(x_{\alpha(\delta)}^{\delta} - x^{\dagger}, u_{\alpha(\delta)}^{\delta} - u^{\dagger}) \rangle \\ & \leq \|(v^{\text{mod}}, v^{\text{obs}})\| \left(\overline{C}_L \left(\frac{\rho}{m} \|A(x_{\alpha(\delta)}^{\delta}, u_{\alpha(\delta)}^{\delta}) - A(x^{\dagger}, u^{\dagger})\|^m \right. \right. \\ & \quad \left. \left. + \frac{1}{o} \|C(u_{\alpha(\delta)}^{\delta}) - C(u^{\dagger})\|^o \right)^{1/t} + LD_{\xi}^{(x_0, u_0)}((x_{\alpha(\delta)}^{\delta}, u_{\alpha(\delta)}^{\delta}), (x^{\dagger}, u^{\dagger})) \right) \end{aligned}$$

hence with $c = 1 - L \|(v^{\text{mod}}, v^{\text{obs}})\| > 0$

$$cD_{\xi}^{(x_0, u_0)}((x_{\alpha(\delta)}^{\delta}, u_{\alpha(\delta)}^{\delta}), (x^{\dagger}, u^{\dagger})) \leq \|(v^{\text{mod}}, v^{\text{obs}})\| \left(\frac{\overline{C}_L(\tau + 1)\delta^o}{o} \right)^{1/t}.$$

In case of (23) we get from minimality (cf. (17))

$$\mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta)) + \alpha \mathcal{R}_2(x_\alpha^\delta) \leq \frac{1}{\alpha} \delta^o + \alpha \mathcal{R}_2(x^\dagger),$$

hence after division by α and by definition (21) of the Bregman distance, as well as the source condition (22)

$$\begin{aligned} & \frac{\mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta))}{\alpha} + D_\xi^{(x_0, u_0)}((x_\alpha^\delta, u_\alpha^\delta), (x^\dagger, u^\dagger)) \\ & \leq \frac{\delta^o}{\alpha} - \langle (v^{\text{mod}}, v^{\text{obs}}), \mathbf{F}'(x^\dagger, u^\dagger)(x_\alpha^\delta - x^\dagger, u_\alpha^\delta - u^\dagger) \rangle \\ & \leq \frac{\delta^o}{\alpha} + \left\| (v^{\text{mod}}, v^{\text{obs}}) \right\| \left(\overline{C}_L \left(\mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta)) \right) \right)^{1/t} \\ & \quad + LD_\xi^{(x_0, u_0)}((x_\alpha^\delta, u_\alpha^\delta), (x^\dagger, u^\dagger)), \end{aligned}$$

hence with c as above and using Young's inequality

$$\begin{aligned} & \frac{\mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta))}{\alpha} + cD_\xi^{(x_0, u_0)}((x_\alpha^\delta, u_\alpha^\delta), (x^\dagger, u^\dagger)) \\ & \leq \frac{\delta^o}{\alpha} + \left\| (v^{\text{mod}}, v^{\text{obs}}) \right\| \overline{C}_L \left(\mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta)) \right)^{1/t} \\ & \leq \frac{\delta^o}{\alpha} + \left(\left(\frac{\left\| (v^{\text{mod}}, v^{\text{obs}}) \right\| \overline{C}_L}{t} \right)^t 2\alpha \right)^{\frac{1}{t-1}} + \frac{\mathcal{S}(\mathbf{F}(x_\alpha^\delta, u_\alpha^\delta), (0, y^\delta))}{2\alpha}, \end{aligned}$$

which by the choice (23) yields (24). \square

REMARK 2. Note that in case of regularization with respect to x only, i.e., (11), we do get boundedness also of the u part via Assumptions 1, 5. However, the sharp estimate (16), that was crucially used in the rates proof above, fails to hold in general. Therefore we do not expect to get (optimal) convergence rates in that case.

2.3. Comparison of implementation. To compare the all-at-once Tikhonov method (9), with (10) and (11) or (12) with Tikhonov regularization for the reduced formulation

$$(25) \quad \min_{x \in \mathcal{D}(F)} \frac{1}{\alpha} \|F(x) - y^\delta\|_Y^o + \frac{\alpha}{r} \|x - x_0\|^r$$

we write the latter as a PDE constrained minimization problem

$$(26) \quad \min_{(x, u) \in \mathcal{D}(\mathbf{F})} \frac{1}{\alpha} \|C(u) - y^\delta\|_Y^o + \frac{\alpha}{r} \|x - x_0\|^r \quad \text{s.t. } A(x, u) = 0,$$

and denote the Tikhonov minimizer as well as its corresponding state by a bar, i.e., \bar{x}_α^δ minimizes (26) and $A(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta) = 0$. For convenience of exposition we will assume W to be reflexive and identify it with its bidual.

We first of all show that reduced Tikhonov regularization (26), (25) is equivalent to all-at-once Tikhonov regularization (9), with (10) and (11) in case $m = 1$ with ρ sufficiently large. This is due to exact penalization, cf., e.g., [18, Theorem 17.3], whose proof remains valid in the Banach space setting. More precisely, for this purpose ρ has to be larger than the norm of the adjoint state, i.e., the solution $\bar{w}_\alpha^\delta \in W$ to

$$A_u(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta)^* w = C'(\bar{u}_\alpha^\delta)^* J_o^Y(C(\bar{u}_\alpha^\delta) - y^\delta) \text{ in } V^*$$

for $(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta)$ solving (26). Existence of this adjoint state is ensured by Fréchet differentiability of C together with Assumption 1, which also implies that

$$(27) \quad \|\bar{w}_\alpha^\delta\|_W \leq C_A \|C'(\bar{u}_\alpha^\delta)\| \|C(\bar{u}_\alpha^\delta) - y^\delta\|_Y^{\sigma-1}.$$

Moreover, by the same arguments as in the proof of Theorem 1, uniform (wrt. $\delta > 0$) boundedness of $(\bar{x}_{\alpha(\delta)}^\delta, \bar{u}_{\alpha(\delta)}^\delta)$ can be concluded from its definition as a minimizer of (25)

$$\frac{1}{o} \frac{\|F(\bar{x}_{\alpha(\delta)}^\delta) - y^\delta\|_Y^o}{\alpha(\delta)} + \frac{1}{r} \|\bar{x}_{\alpha(\delta)}^\delta - x_0\|^r \leq \frac{1}{o} \frac{\delta^o}{\alpha(\delta)} + \frac{1}{r} \|x^\dagger - x_o\|^r$$

with $\alpha(\delta)$ chosen a priori according to (13) or a posteriori according to

$$\delta^o \leq \|F(\bar{x}_{\alpha(\delta)}^\delta) - y^\delta\|_Y^o \leq \tau \delta^o,$$

(which due to exact penalization will finally coincide with (14),) together with the identity (cf. (18))

$$\begin{aligned} 0 &= A(\bar{x}_{\alpha(\delta)}^\delta, \bar{u}_{\alpha(\delta)}^\delta) = A(x_\alpha^\delta, u_\alpha^\delta) \\ &= \int_0^1 \left(A_x(x^\theta, u^\theta)(\bar{x}_{\alpha(\delta)}^\delta - x^\dagger) + A_u(x^\theta, u^\theta)(\bar{u}_{\alpha(\delta)}^\delta - u^\dagger) \right) d\theta \end{aligned}$$

with $x^\theta = x^\dagger + \theta(\bar{x}_{\alpha(\delta)}^\delta - x^\dagger)$, $u^\theta = u^\dagger + \theta(\bar{u}_{\alpha(\delta)}^\delta - u^\dagger)$, which yields an estimate like (19) with $\bar{C}_\delta = 0$ and $(x_\alpha^\delta, u_\alpha^\delta)$ replaced by $(\bar{x}_{\alpha(\delta)}^\delta, \bar{u}_{\alpha(\delta)}^\delta)$. Thus, from (27) we get a uniform bound on $\|\bar{w}_\alpha^\delta\|_W$ and can conclude the following equivalence.

PROPOSITION 3. *Let Assumptions 1, 5 be satisfied and let C be Fréchet differentiable with C' mapping bounded sets to bounded sets. There exist $\rho > 0$ sufficiently large and $\bar{\delta} > 0$ sufficiently small such that for all $\delta \in (0, \bar{\delta})$ and $\alpha(\delta)$ chosen according to (13) or (14), the minimizers of the all-at-once Tikhonov functional according to (9) with (10), (11) with $m = 1$ coincide with those of the reduced Tikhonov functional (25).*

We now consider first order optimality conditions for the reduced and the all-at-once formulations (26), (9) with general $m \geq 1$, so that they are not necessarily equivalent, and for this purpose assume A , C , and the occurring norms to be continuously Fréchet differentiable (i.e., the corresponding spaces to be uniformly smooth). In case $\mathcal{D}(\mathbf{F}) = X \times V$, we get the first order necessary conditions

$$(28) \quad \begin{aligned} A(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta) &= 0 \\ \alpha J_r^X(\bar{x}_\alpha^\delta - x_0) &= -A_x(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta)^* \bar{w}_\alpha^\delta \\ A_u(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta)^* \bar{w}_\alpha^\delta &= -C'(\bar{u}_\alpha^\delta)^* J_o^Y(C(\bar{u}_\alpha^\delta) - y^\delta) \end{aligned}$$

for a minimizer $(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta)$ of the reduced Tikhonov functional (26) and

$$\begin{aligned} \alpha J_r^X(x_\alpha^\delta - x_0) &= -\rho A_x(x_\alpha^\delta, u_\alpha^\delta)^* J_m^{W*}(A(x_\alpha^\delta, u_\alpha^\delta)) \\ \rho A_u(x_\alpha^\delta, u_\alpha^\delta)^* J_m^{W*}(A(x_\alpha^\delta, u_\alpha^\delta)) &= -C'(\bar{u}_\alpha^\delta)^* J_o^Y(C(\bar{u}_\alpha^\delta) - y^\delta) [-\alpha J_r^V(u_\alpha^\delta - u_0)] \end{aligned}$$

for a minimizer $(x_\alpha^\delta, u_\alpha^\delta)$ of the all-at-once Tikhonov functional (9) with (10) and (11) or (12), where the term in brackets is skipped in case of (11). Upon defining w_α^δ as the

solution to $A_u(x_\alpha^\delta, u_\alpha^\delta)^* w = -C'(u_\alpha^\delta)^* J_o^Y(C(u_\alpha^\delta) - y^\delta) [-\alpha J_r^V(u_\alpha^\delta - u_0)]$, as justified by Assumption 1, the latter can be rewritten as

$$(29) \quad \begin{aligned} \rho J_m^{W*}(A(x_\alpha^\delta, u_\alpha^\delta)) &= w_\alpha^\delta \\ \alpha J_r^X(x_\alpha^\delta - x_0) &= -A_x(x_\alpha^\delta, u_\alpha^\delta)^* w_\alpha^\delta \\ A_u(x_\alpha^\delta, u_\alpha^\delta)^* w_\alpha^\delta &= -C'(u_\alpha^\delta)^* J_o^Y(C(u_\alpha^\delta) - y^\delta) [-\alpha J_r^V(u_\alpha^\delta - u_0)] \end{aligned}$$

so very similar to (28). These first order necessary optimality conditions can be justified, e.g., by a Zowe Kurcyusz constraint qualification [22, Section 6.1], which in the all-at-once case is an empty condition and in the reduced case amounts to surjectivity of $A'(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta)$ and is thus obviously satisfied, e.g., under Assumption 1.

In case of additional convex constraints on the parameters $\mathcal{D}(\mathbf{F}) = \mathcal{C} \times V$ with some convex set \mathcal{C} , e.g., defined by the pointwise bounds \underline{a} , \bar{a} , \underline{c} , \bar{c} in (3), (4), the second line in (28), (29) changes to

$$(30) \quad \langle \alpha J_r^X(\tilde{x} - x_0) + A_x(\tilde{x}, \tilde{u})^* \tilde{w}, x - \tilde{x} \rangle_{X^*, X} \geq 0 \text{ for all } x \in \mathcal{C}$$

for $(\tilde{x}, \tilde{u}, \tilde{w}) = (\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta, \bar{w}_\alpha^\delta)$, $(\tilde{x}, \tilde{u}, \tilde{w}) = (x_\alpha^\delta, u_\alpha^\delta, w_\alpha^\delta)$, respectively, and the Zowe Kurcyusz constraint qualification is again always satisfied in the all-at-once case and amounts to

$$A_x(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta) \mathcal{C}(\bar{x}_\alpha^\delta) + A_u(\bar{x}_\alpha^\delta, \bar{u}_\alpha^\delta) V = W^*$$

with

$$\mathcal{C}(\bar{x}_\alpha^\delta) = \{\gamma(x - \bar{x}_\alpha^\delta) : \gamma \geq 0, x \in \mathcal{C}\}$$

in the reduced case, which again obviously holds under Assumption 1.

3. The iteratively regularized Gauss-Newton method. Throughout this section we assume A and C to be continuously Fréchet differentiable on $\mathcal{D}(\mathbf{F})$ and abbreviate, analogously to (10)

$$(31) \quad \begin{aligned} &\mathcal{S}(\mathbf{F}(\tilde{x}, \tilde{u}) + \mathbf{F}'(\tilde{x}, \tilde{u})(x - \tilde{x}, u - \tilde{u}), (y^{\text{mod}}, y^{\text{obs}})) \\ &= \frac{\rho}{m} \left\| A(\tilde{x}, \tilde{u}) + A_x(\tilde{x}, \tilde{u})(x - \tilde{x}) + A_u(\tilde{x}, \tilde{u})(u - \tilde{u}) - y^{\text{mod}} \right\|_{W^*}^m \\ &\quad + \frac{1}{o} \left\| C(\tilde{u}) + C'(\tilde{u})(u - \tilde{u}) - y^{\text{obs}} \right\|_Y^o. \end{aligned}$$

Given some iterate $(x_k^\delta, u_k^\delta) \in \mathcal{D}(\mathbf{F})$, we define the next Newton (more, precisely, iteratively regularized Gauss Newton IRGNM, cf., e.g. [1, 2, 15, 13, 16, 11, 26]) iterate $(x_{k+1}^\delta(\alpha), u_{k+1}^\delta(\alpha))$ as a minimizer of

$$(32) \quad \min_{(x, u) \in \mathcal{D}(\mathbf{F})} \mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta) + \mathbf{F}'(x_k^\delta, u_k^\delta)(x - x_k^\delta, u - u_k^\delta), (0, y^\delta)) + \alpha \mathcal{R}(x, u)$$

with

$$(33) \quad \mathcal{R}(x, u) = \mathcal{R}_1(x) = \frac{1}{r} \|x - x_0\|_X^r$$

or

$$(34) \quad \mathcal{R}(x, u) = \mathcal{R}_2(x, u) = \frac{1}{r} \|x - x_0\|_V^r + \frac{1}{q} \|u - u_0\|_V^q,$$

like in the previous section, cf. (11), (12).

3.1. Well-definedness and convergence. As compared to Tikhonov regularization, iterative regularization methods for nonlinear problems can only be proven to converge under certain structural assumptions restricting the nonlinearity of the forward operator. We here first of all concentrate on convergence under the simple tangential cone condition.

ASSUMPTION 7.

$$\mathcal{S}(\mathbf{F}(\tilde{x}, \tilde{u}) + \mathbf{F}'(\tilde{x}, \tilde{u})(x - \tilde{x}, u - \tilde{u}), \mathbf{F}(x)) \leq c_{tc} \mathcal{S}(\mathbf{F}(x, u), \mathbf{F}(\tilde{x}, \tilde{u}))$$

for some $0 < c_{tc} < 1$ and all $(x, u), (\tilde{x}, \tilde{u}) \in \mathcal{D}(\mathbf{F}) \cap \mathcal{B}_\varrho(x_0, u_0)$ with $\mathcal{B}_\varrho(x_0, u_0)$ a closed ball with sufficiently small radius $\varrho > 0$ around (x_0, u_0) .

Correspondingly, we consider a posteriori choice of the stopping index $k_* = k_*(\delta)$ by the discrepancy principle

$$(35) \quad k_* = \min\{k \in \mathbb{N} : \mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta), (0, y^\delta)) \leq \frac{\tau \delta^o}{o}\}$$

and of α_k for $k \leq k_*$ by the inexact Newton strategy

$$(36) \quad \underline{\sigma} \leq \sigma_k(x_{k+1}^\delta(\alpha_k), u_{k+1}^\delta(\alpha_k)) \leq \bar{\sigma}$$

where

$$\sigma_k(x, u) = \frac{\mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta) + \mathbf{F}'(x_k^\delta, u_k^\delta)(x - x_k^\delta, u - u_k^\delta), (0, y^\delta))}{\mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta), (0, y^\delta))}$$

is the ratio between the predicted and the old data misfit. Note that by definition (35) of k_* , the denominator in $\sigma_k(x_{k+1}^\delta(\alpha_k), u_{k+1}^\delta(\alpha_k))$ is nonzero for $k < k_*$ and $\delta > 0$.

PROPOSITION 4. Let $X \times V$ be reflexive and uniformly convex and let Assumptions 2, 4, 7 be satisfied with c_{tc} sufficiently small

$$c_{tc} < \underline{\sigma} < \bar{\sigma} < 1$$

and either (i) $\mathbf{F}'(x, u)$ be weakly closed for all $(x, u) \in \mathcal{D}(\mathbf{F})$ and Y be reflexive or (ii) $\mathcal{D}(\mathbf{F})$ be weakly closed. In case of regularization with respect to x only (33), we additionally impose Assumptions 1, 5.

Moreover, let τ be chosen sufficiently large so that

$$(37) \quad C_S(c_{tc} + \frac{1 + C_S c_{tc}}{\tau}) \leq \underline{\sigma} \text{ and } c_{tc} < \frac{1 - \bar{\sigma}}{2},$$

for

$$C_S = \max\left\{\frac{\rho}{m} 2^{m-1}, \frac{1}{o} 2^{o-1}\right\},$$

let (x_0, u_0) be close enough to (x^\dagger, u^\dagger) with respect to the Bregman distance, and let the signal-to-noise ratio condition

$$\tau < \frac{o \mathcal{S}(\mathbf{F}(x_0, u_0), (0, y^\delta))}{\delta^o}$$

hold.

- Then for all $k \leq k_*(\delta) - 1$ with $k_*(\delta)$ according to (35), the iterates

$$(x_{k+1}^\delta, u_{k+1}^\delta) := \begin{cases} (x_{k+1}^\delta(\alpha_k), u_{k+1}^\delta(\alpha_k)), & \alpha_k \text{ as in (36)} \\ (x_0, u_0) & \text{else} \end{cases} \quad \text{if } \sigma_k(x_0, u_0) \geq \bar{\sigma}$$

and the stopping index $k_*(\delta)$ are well-defined by (32) with (31) and (33) or (34).

- There exists $C > 0$ independent of $\delta \in (0, \bar{\delta})$ such that

$$(38) \quad \forall \delta \in (0, \bar{\delta}) : \left\| x_{k_*}^\delta - x_0 \right\|_X^r \leq \left\| x^\dagger - x_0 \right\|_X^r \text{ and } \left\| u_{k_*}^\delta \right\|_V \leq C$$

in case of minimizers of (32) with (31) and (33) and

$$(39) \quad \forall \delta \in (0, \bar{\delta}) : \mathcal{R}_2(x_{k_*}^\delta, u_{k_*}^\delta) \leq \mathcal{R}_2(x^\dagger, u^\dagger)$$

in case of minimizers of (32) with (31) and (34).

- In both cases (38), (39), the regularized approximations $(x_{k_*}^\delta, u_{k_*}^\delta)$ converges weakly subsequentially to a solution of (1), (2) as $\delta \rightarrow 0$.

Proof. See [16, Theorem 3].

It only remains to prove boundedness of the u part in case (33). Again we concentrate on the convergence part of the proof and use minimality and the tangential cone condition Assumption 7 to obtain

$$\begin{aligned} & \mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta) + \mathbf{F}'(x_k^\delta, u_k^\delta)(x_{k+1}^\delta - x_k^\delta, u_{k+1}^\delta - u_k^\delta), (0, y^\delta)) + \frac{\alpha_k}{r} \left\| x_{k+1}^\delta - x_0 \right\|_X^r \\ & \leq \mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta) + \mathbf{F}'(x_k^\delta, u_k^\delta)(x^\dagger - x_k^\delta, u^\dagger - u_k^\delta), (0, y^\delta)) + \frac{\alpha_k}{r} \left\| x^\dagger - x_0 \right\|_X^r \\ & \leq C_S(1 + C_S c_{tc})\delta + C_S c_{tc} \mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta), (0, y^\delta)) + \frac{\alpha_k}{r} \left\| x^\dagger - x_0 \right\|_X^r, \end{aligned}$$

which for $k \leq k_* - 1$ by (35), (36), yields

$$\begin{aligned} & \left(\underline{\sigma} - C_S(c_{tc} + \frac{1 + C_S c_{tc}}{\tau}) \right) \mathcal{S}(\mathbf{F}(x_k^\delta, u_k^\delta), (0, y^\delta)) + \frac{\alpha_k}{r} \left\| x_{k+1}^\delta - x_0 \right\|_V^r \\ & \leq \frac{\alpha_k}{r} \left\| x^\dagger - x_0 \right\|_V^r, \end{aligned}$$

which upon division by α_k gives a uniform bound on $\left\| x_{k+1}^\delta - x_0 \right\|_V$ (namely the one stated in (38); obviously, (39) can be obtained in the same manner). Thus up to the positive factor $\left(\underline{\sigma} - C_S(c_{tc} + \frac{1 + C_S c_{tc}}{\tau}) \right)$ we are in the same situation as in (17) and thus can use the same arguments to prove uniform boundedness of $u_{k_*}^\delta$. As a matter of fact, since we only consider the a posteriori regularization parameter choice now, the situation is even simpler, since we can use the upper estimate in (35) in the identity (18) (with $(x_\alpha^\delta, u_\alpha^\delta)$ replaced by (x_k^δ, u_k^δ)). \square

A sufficient condition for Assumption 7 to hold is, like in the reduced case, the adjoint range invariance condition

$$(40) \quad \mathbf{F}'(\tilde{x}, \tilde{u}) = \mathbf{R}_{(x,u)}^{(\tilde{x}, \tilde{u})} \mathbf{F}'(x, u) \text{ with } \left\| \mathbf{R}_{(x,u)}^{(\tilde{x}, \tilde{u})} - I \right\| \leq c_R < 1$$

for some $0 < c_{tc} < 1$ and all $(x, u), (\tilde{x}, \tilde{u}) \in \mathcal{D}(\mathbf{F}) \cap \mathcal{B}_\rho(x_0, u_0)$, as well as linear operators $\mathbf{R}_{(x,u)}^{(\tilde{x}, \tilde{u})} : W^* \times Y \rightarrow W^* \times Y$. More explicitly, with $\mathbf{R}_{(x,u)}^{(\tilde{x}, \tilde{u})} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$ the identity between the derivatives of \mathbf{F} can be rewritten as

$$(41) \quad \begin{aligned} A_x(\tilde{x}, \tilde{u}) &= R_{11} A_x(x, u) \\ A_u(\tilde{x}, \tilde{u}) &= R_{11} A_u(x, u) + R_{12} C \\ 0 &= R_{21} A_x(x, u) \\ C'(\tilde{u}) &= R_{21} A_u(x, u) + R_{22} C'(u) \end{aligned}$$

Actually, this is the condition that we will check for some of our examples.

Closely related is the following range invariance condition

ASSUMPTION 8.

$$(42) \quad \mathbf{F}'(\tilde{x}, \tilde{u}) = \mathbf{F}'(x, u) \mathbf{R}_{(x, u)}^{(\tilde{x}, \tilde{u})} \text{ with } \|\mathbf{R}_{(x, u)}^{(\tilde{x}, \tilde{u})} - I\| \leq c_R < 1$$

for some $0 < c_{tc} < 1$ and all $(x, u), (\tilde{x}, \tilde{u}) \in \mathcal{D}(\mathbf{F}) \cap \mathcal{B}_\rho(x_0, u_0)$, as well as linear operators $\mathbf{R}_{(x, u)}^{(\tilde{x}, \tilde{u})} : X \times V \rightarrow X \times V$.

Again we can write the identity between the derivatives of \mathbf{F} more explicitly by using the block decomposition $\mathbf{R}_{(x, u)}^{(\tilde{x}, \tilde{u})} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$:

$$(43) \quad \begin{aligned} A_x(\tilde{x}, \tilde{u}) &= A_x(x, u)R_{11} + A_u(x, u)R_{21} \\ A_u(\tilde{x}, \tilde{u}) &= A_x(x, u)R_{12} + A_u(x, u)R_{22} \\ 0 &= C'(u)R_{21} \\ C'(\tilde{u}) &= C'(u)R_{22} \end{aligned}$$

Under this alternative condition we get convergence with an a priori choice of α_k and k_* , however only in Hilbert space. Note that convergence under such a range invariance condition in the general Banach space setting is still an open problem also for the conventional reduced formulation. We quote a convergence result for the case of (34).

PROPOSITION 5. *Let $o = m = r = 2$ and $X \times V$ and $W^* \times V$ be Hilbert spaces and let Assumptions 2, 4, 8 be satisfied with c_{tc} sufficiently small. Moreover, let τ be chosen sufficiently large, and let (x_0, u_0) be close enough to (x^\dagger, u^\dagger) .*

Then for all $k \in \mathbb{N}$, the iterates $(x_{k+1}^\delta, u_{k+1}^\delta) = (x_{k+1}^\delta(\alpha_k), u_{k+1}^\delta(\alpha_k))$ with an apriori chosen sequence $(\alpha_k)_{k \in \mathbb{N}}$ satisfying

$$(44) \quad \alpha_k \rightarrow 0 \text{ as } k \rightarrow \infty \text{ and } 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq C_\alpha$$

for some $C_\alpha \geq 1$ are well-defined by (32) with (31) and (34), and with a choice of $k_(\delta)$ such that*

$$(45) \quad k_*(\delta) \rightarrow \infty \text{ and } \frac{\delta}{\sqrt{\alpha_{k_*(\delta)}}} \rightarrow 0 \text{ as } \delta \rightarrow 0$$

$(x_{k_(\delta)}^\delta, u_{k_*(\delta)}^\delta)$ converges strongly to a solution of (1), (2) as $\delta \rightarrow 0$.*

Proof. See [12, Theorem 2.7]. □

3.2. Convergence rates. Exactly as in the proof of the first part (case of a posteriori regularization parameter choice) of Proposition 2, using (39) we get rates under a source condition (cf. [13, 26] for the reduced case).

PROPOSITION 6. *Let \mathbf{F}' satisfy Assumption 6 and let, for and some $(v^{mod}, v^{obs}) \in W \times Y^*$ with $L \|(v^{mod}, v^{obs})\| < 1$ the source condition (22) hold.*

Then for the Newton iterates according to (34) with the a posteriori choice (35), (36) or the a priori choice (44), $\alpha_{k^}(\delta) \sim \delta^{\frac{o}{t^*}}$ (cf. (23)), we have*

$$D_\xi^{(x_0, u_0)}((x_{k^*}^\delta, u_{k^*}^\delta), (x^\dagger, u^\dagger)) = O(\delta^{o/t})$$

(cf. (24)).

3.3. Comparison of implementation. We compare one step of the the all-at-once IRGNM to one step of the reduced IRGNM

$$(46) \quad \min_{x \in \mathcal{D}(F)} \frac{1}{o} \|F(\bar{x}_k^\delta) + F'(\bar{x}_k^\delta)(x - \bar{x}_k^\delta) - y^\delta\|_Y^o + \frac{\alpha}{r} \|x - x_0\|_V^r,$$

where $\tilde{\mathcal{D}} = \{(x, u, \tilde{u}) : (\bar{x}_k^\delta, \tilde{u}) \in \mathcal{D}(\mathbf{F}), x \in X, u \in V\}$, and denote the iterates resulting from the reduced version by a bar. This can be rewritten as a PDE constrained minimization problem (setting $\tilde{u} = S(\bar{x}_k^\delta)$, $u = \tilde{u} + S'(\bar{x}_k^\delta)(x - \bar{x}_k^\delta) = \tilde{u} - A_u(\bar{x}_k^\delta, \tilde{u})^{-1} A_x(\bar{x}_k^\delta, \tilde{u})(x - \bar{x}_k^\delta)$)

$$(47) \quad \begin{aligned} & \min_{(x, u, \tilde{u}) \in \tilde{\mathcal{D}}} \frac{1}{o} \|C(\tilde{u}) + C'(\tilde{u})(u - \tilde{u}) - y^\delta\|_Y^o + \frac{\alpha}{r} \|x - x_0\|^r \\ & \text{s.t.} \quad \begin{cases} A(\bar{x}_k^\delta, \tilde{u}) = 0, \\ A(\bar{x}_k^\delta, \tilde{u}) + A_x(\bar{x}_k^\delta, \tilde{u})(x - \bar{x}_k^\delta) + A_u(\bar{x}_k^\delta, \tilde{u})(u - \tilde{u}) = 0, \end{cases} \end{aligned}$$

in particular it contains the nonlinear model $A(\bar{x}_k^\delta, \tilde{u}) = 0$ as well as the linearized one as constraints.

Again we assume W to be reflexive and start with the exact penalization case $m = 1$ in (32). However, the PDE constrained minimization problem to which (32) with $m = 1$ and ρ sufficiently large is equivalent (analogously to Proposition 3) is the following

$$(48) \quad \begin{aligned} & \min_{(x, u, \tilde{u}) \in \tilde{\mathcal{D}}} \frac{1}{o} \|C(u_k^\delta) + C'(u_k^\delta)(u - u_k^\delta) - y^\delta\|_Y^o + \frac{\alpha}{r} \|x - x_0\|^r \\ & \text{s.t.} \quad A(x_k^\delta, u_k^\delta) + A_x(x_k^\delta, u_k^\delta)(x - x_k^\delta) + A_u(x_k^\delta, u_k^\delta)(u - u_k^\delta) = 0 \end{aligned}$$

so only the linearized model appears as a constraint here, and the problem is obviously not the same as the reduced one (47), differently from Section 2. For the sake of completeness, we point out that exact penalization indeed holds for sufficiently large but finite ρ since the adjoint state w_k for (48) by Assumption 1 and Proposition 4 is well defined by

$$A_u(x_k^\delta, u_k^\delta)^* w_k = -C'(u_k^\delta)^* J_o^Y(C(u_k^\delta) + C'(u_k^\delta)(u_{k+1}^\delta - u_k^\delta) - y^\delta) \text{ in } V^*$$

and uniformly (wrt $\delta \in (0, \bar{\delta})$ and $k \leq k_*(\delta)$) bounded, thus it is possible to choose such a uniform penalty parameter $\rho \geq \sup_{\delta \in (0, \bar{\delta})} \max_{k \in \{1, \dots, k_*(\delta)\}} \|w_k\|_W$.

The first order optimality conditions for the reduced and the all-at-once formulations (47), (32), (48) in case $\mathcal{D}(\mathbf{F}) = X \times V$ read as follows:

$$(49) \quad \begin{aligned} & A(\bar{x}_k^\delta, \bar{u}_k^\delta) = 0 \\ & A(\bar{x}_k^\delta, \bar{u}_k^\delta) + A_x(\bar{x}_k^\delta, \bar{u}_k^\delta)(\bar{x}_{k+1}^\delta - \bar{x}_k^\delta) + A_u(\bar{x}_k^\delta, \bar{u}_k^\delta)(\bar{u}_{k+1}^\delta - \bar{u}_k^\delta) = 0 \\ & \alpha J_r^X(\bar{x}_{k+1}^\delta - x_0) = -A_x(\bar{x}_k^\delta, \bar{u}_k^\delta)^* \bar{w}_{k+1}^\delta \\ & A_u(\bar{x}_k^\delta, \bar{u}_k^\delta)^* \bar{z}_k^\delta = -(C''(\bar{u}_k^\delta)(\bar{u}_{k+1}^\delta - \bar{u}_k^\delta))^* J_o^Y(C(\bar{u}_k^\delta) + C'(\bar{u}_k^\delta)(\bar{u}_{k+1}^\delta - \bar{u}_k^\delta) - y^\delta) \\ & \quad - (A_{ux}(\bar{x}_k^\delta, \bar{u}_k^\delta)(\bar{x}_{k+1}^\delta - \bar{x}_k^\delta) + A_{uu}(\bar{x}_k^\delta, \bar{u}_k^\delta)(\bar{u}_{k+1}^\delta - \bar{u}_k^\delta))^* \bar{w}_{k+1}^\delta \\ & A_u(\bar{x}_k^\delta, \bar{u}_k^\delta)^* \bar{w}_{k+1}^\delta = -C'(\bar{u}_k^\delta)^* J_o^Y(C(\bar{u}_k^\delta) + C'(\bar{u}_k^\delta)(\bar{u}_{k+1}^\delta - \bar{u}_k^\delta) - y^\delta) \end{aligned}$$

for a minimizer $(\bar{x}_{k+1}^\delta, \bar{u}_k^\delta, \bar{u}_{k+1}^\delta)$ of the reduced version (47), (where the equation for the Lagrange multiplier \bar{z}_k^δ corresponding to the second PDE constraint may be

skipped)

$$\begin{aligned}
(50) \quad & \rho J_m^{W*}(A(x_k^\delta, u_k^\delta) + A_x(x_k^\delta, u_k^\delta)(x_{k+1}^\delta - x_k^\delta) + A_u(x_k^\delta, u_k^\delta)(u_{k+1}^\delta - u_k^\delta)) = w_{k+1}^\delta \\
& \alpha J_r^X(x_{k+1}^\delta - x_0) = -A_x(x_k^\delta, u_k^\delta)^* w_{k+1}^\delta \\
& A_u(x_k^\delta, u_k^\delta)^* w_{k+1}^\delta = -C'(u_k^\delta)^* J_o^Y(C(u_k^\delta) + C'(u_k^\delta)(u_{k+1}^\delta - u_k^\delta) - y^\delta)
\end{aligned}$$

for a minimizer $(x_{k+1}^\delta, u_{k+1}^\delta)$ of the all-at-once version (32), and

$$\begin{aligned}
(51) \quad & A(x_k^\delta, u_k^\delta) + A_x(x_k^\delta, u_k^\delta)(x_{k+1}^\delta - x_k^\delta) + A_u(x_k^\delta, u_k^\delta)(u_{k+1}^\delta - u_k^\delta) = 0 \\
& \alpha J_r^X(x_{k+1}^\delta - x_0) = -A_x(x_k^\delta, u_k^\delta)^* w_{k+1}^\delta \\
& A_u(x_k^\delta, u_k^\delta)^* w_{k+1}^\delta = -C'(u_k^\delta)^* J_o^Y(C(u_k^\delta) + C'(u_k^\delta)(u_{k+1}^\delta - u_k^\delta) - y^\delta)
\end{aligned}$$

for a minimizer $(x_{k+1}^\delta, u_{k+1}^\delta)$ of the all-at-once version (32) in case $m = 1$ and ρ sufficiently large, i.e., of (48).

Justification of these first order optimality conditions by Fréchet differentiability of A and C and a Zowe Kurcyusz constraint qualification is again trivial in case of (32) with $m > 1$. In case of (47), we need surjectivity of

$$G'(x, u, \tilde{u}) = \begin{pmatrix} 0 & 0 & A_u(\bar{x}_k^\delta, \tilde{u}) \\ A_x(\bar{x}_k^\delta, \tilde{u}) & A_u(\bar{x}_k^\delta, \tilde{u}) & A_{xu}(\bar{x}_k^\delta, \tilde{u})(x - \bar{x}_k^\delta) + A_{xu}(\bar{x}_k^\delta, \tilde{u})(u - \tilde{u}) \end{pmatrix}$$

at $(x, u, \tilde{u}) = (\bar{x}_{k+1}^\delta, \bar{u}_{k+1}^\delta, \bar{u}_k^\delta)$ and in case of (48) surjectivity of

$$G'(x, u) = (A_x(x_k^\delta, u_k^\delta) \quad A_u(x_k^\delta, u_k^\delta))$$

at $(x, u) = (x_{k+1}^\delta, u_{k+1}^\delta)$, so that in both cases the constraint qualifications are satisfied under Assumption 1.

In case of additional convex constraints on the parameters $\mathcal{D}(\mathbf{F}) = \mathcal{C} \times V$ with some convex set \mathcal{C} , the second lines in (49), (50), (49) again change to variational inequalities of the form (30).

4. Landweber iteration. Gradient methods applied to the unregularized least squares problem (i.e., (25) or (9) with $\alpha = 0$) lead to the Landweber iteration in a reduced

$$\begin{aligned}
(52) \quad & \bar{x}_{k+1}^{\delta*} = \bar{x}_k^{\delta*} - \mu_k F'(\bar{x}_k^{\delta*})^* J_o^Y(F(\bar{x}_k^{\delta*}) - y^\delta) \\
& \bar{x}_{k+1}^\delta = J_{r^*}^{X*}(\bar{x}_{k+1}^{\delta*})
\end{aligned}$$

and an all-at-once setting

$$\begin{aligned}
(53) \quad & (x_{k+1}^{\delta*}, u_{k+1}^{\delta*}) = (x_k^{\delta*}, u_k^{\delta*}) - \mu_k \mathbf{F}'(x_k^{\delta*}, u_k^{\delta*})^* (J_m^{W*}, J_o^Y)(\mathbf{F}(x_k^{\delta*}, u_k^{\delta*}) - (0, y^\delta)) \\
& (x_{k+1}^\delta, u_{k+1}^\delta) = (J_{r^*}^{X*}(x_{k+1}^{\delta*}), J_{q^*}^{V*}(u_{k+1}^{\delta*}))
\end{aligned}$$

with appropriately chosen stepsizes μ_k , cf., e.g., [16]. We here assume that X and V are smooth and s -convex so that the duality mappings $J_{r^*}^{X*}$, $J_{q^*}^{V*}$ are single valued and in fact inverses of J_r^X , J_q^V , respectively. Moreover, we postulate continuous Fréchet differentiability of A and C , and, for (52), also Assumption 1 to guarantee well-definedness of F .

Under a tangential cone condition

$$\|F(\tilde{x}) + F'(\tilde{x})(x - \tilde{x}) - F(x)\|_Y \leq c_{tc} \|F(\tilde{x}) - F(x)\|_Y$$

for some $0 < c_{tc} < 1$ and all $x \in \mathcal{B}_\rho(x_0)$ in the reduced case (52) or Assumption 7 in the all-at-once case (53), we have convergence of the iterates \bar{x}_k^0 or (x_k^0, u_k^0) to a solution of (1), (2) as $k \rightarrow \infty$ in the exact case $\delta = 0$. Under the same conditions, with additionally Y or $W^* \times Y$ being uniformly smooth, an a posteriori stopping rule according to

$$k_* = \min\{k \in \mathbb{N} : \|F(\bar{x}_k^\delta), y^\delta\|_Y \leq \tau\delta\},$$

for (52) or (35) for (53) yields convergence of $\bar{x}_{k_*}^\delta$ or $(x_{k_*}^\delta, u_{k_*}^\delta)$ to a solution of (5) or (1), (2) as $\delta \rightarrow 0$ holds. This is an immediate consequence of Theorems 1 and 2 in [16]. Note that even if the all-at-once tangential cone condition Assumption 7 might be hard to verify for nonlinear problems, Landweber still makes sense for linear problems where this condition is trivially satisfied. And it actually makes a big difference to the reduced version also in the linear case as we will see in the following.

4.1. Comparison of implementation. Although we do not deal with minimization problems now, each reduced Landweber step can be split into a system that resembles (28), but is has triangular (so not fully coupled) structure. Namely, using the definition of $F = C \circ S$, the chain rule, and the Implicit Function Theorem for differentiating S

$$A(x, S(x)) = 0, \quad A_x(x, S(x)) + A_u(x, S(x))S'(x) = 0,$$

i.e.,

$$F'(x)^* J_o^Y(F(x) - y^\delta) = -A_x(x, S(x))^* (A_u(x, S(x))^*)^{-1} C'(S(x))^* J_o^Y(C(S(x)) - y^\delta)$$

and setting $\bar{u}_k^\delta = S(\bar{x}_k^\delta)$, $\bar{w}_k^\delta = -(A_u(\bar{x}_k^\delta, \bar{u}_k^\delta))^* C'(\bar{u}_k^\delta)^* J_o^Y(C(\bar{u}_k^\delta) - y^\delta)$ we can rewrite one reduced Landweber step as

$$\begin{aligned} A(\bar{x}_k^\delta, \bar{u}_k^\delta) &= 0 \\ A_u(\bar{x}_k^\delta, \bar{u}_k^\delta)^* \bar{w}_k^\delta &= -C'(\bar{u}_k^\delta)^* J_o^Y(C(\bar{u}_k^\delta) - y^\delta) \\ \bar{x}_{k+1}^{\delta*} &= \bar{x}_k^{\delta*} - \mu_k A_x(\bar{x}_k^\delta, \bar{u}_k^\delta)^* \bar{w}_k^\delta \\ \bar{x}_{k+1}^\delta &= J_{r^*}^{X^*}(\bar{x}_{k+1}^{\delta*}) \end{aligned}$$

which involves solution of a nonlinear and a linearized (adjoint) model, whereas the all-at-once version using

$$\mathbf{F}'(x, u)^* (J_m^{W^*}, J_o^Y)(\mathbf{F}(x, u) - (0, y^\delta)) = \begin{pmatrix} A_x(x, u) & A_u(x, u) \\ 0 & C'(u) \end{pmatrix}^* \begin{pmatrix} J_m^{W^*}(A(x, u)) \\ J_o^Y(C(u) - y^\delta) \end{pmatrix}$$

reads

$$\begin{aligned} x_{k+1}^{\delta*} &= x_k^{\delta*} - \mu_k A_x(x_k^{\delta*}, u_k^{\delta*})^* J_m^{W^*}(A(x_k^{\delta*}, u_k^{\delta*})) \\ u_{k+1}^{\delta*} &= u_k^{\delta*} - \mu_k \left(A_u(x_k^{\delta*}, u_k^{\delta*})^* J_m^{W^*}(A(x_k^{\delta*}, u_k^{\delta*})) + C'(u)^* J_o^Y(C(u) - y^\delta) \right) \\ (x_{k+1}^\delta, u_{k+1}^\delta) &= (J_{r^*}^{X^*}(x_{k+1}^{\delta*}), J_{q^*}^{V^*}(u_{k+1}^{\delta*})), \end{aligned}$$

(where again we have assumed W to be reflexive). So as opposed to all other schemes considered here and to the reduced Landweber setting, in an all-at-once Landweber step no linear or nonlinear model is solved.

5. Application and Numerical Tests. Since the main difference in implementation seems to arise in the context of nonlinear PDEs (at least as far as the IRGNM is concerned) we consider an inverse source problem for a nonlinear PDE, namely, identification of b in

$$(54) \quad -\Delta u + \xi u^3 = b \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega$$

from measurements of u in $\Omega \subseteq \mathbb{R}^d$, $d \in \{1, 2, 3\}$. Here $\xi \in \mathbb{R}$ is a known parameter, which we use in order to study different strengths of nonlinearity of the PDE and also the non-elliptic case arising when ξ is smaller than the negative of the embedding constant $H_0^1(\Omega) \rightarrow L^4(\Omega)$, so that the parameter-to-state map fails to be well-defined and therefore the reduced approach is not applicable. This example has already been used in [14] to compare reduced Tikhonov regularization with an all-at-once version of the IRGNM, both with adaptive discretization.

5.1. Verification of convergence conditions. Indeed, even a more general version of this example satisfies the structural conditions on the forward operator (41), (43) for the all-at-once IRGNM, and also their reduced versions.

LEMMA 7. *Let A, C in (1), (2) have the form $A(x, u) = \hat{A}(u) + Lx$, $C(u) = Cu$ with $L : X \rightarrow W^*$, $C : V \rightarrow Y$ linear operators, $\hat{A} : V \rightarrow W^*$ possibly nonlinear and Fréchet differentiable, and let C^\dagger, L^\dagger denote the Moore Penrose generalized inverse of C and L , respectively.*

(a) *If $\mathcal{N}(\hat{A}'(\tilde{u}) - \hat{A}'(u))^\perp \subseteq \mathcal{N}(C)^\perp$, then (41) is satisfied with*

$$\|\mathbf{R}_{(x,u)}^{(\tilde{x}, \tilde{u})} - I\| = \|(\hat{A}'(\tilde{u}) - \hat{A}'(u))C^\dagger\|$$

(b) *If $\mathcal{R}(\hat{A}'(\tilde{u}) - \hat{A}'(u)) \subseteq \overline{\mathcal{R}(L)}$, then (43) is satisfied with*

$$\|\mathbf{R}_{(x,u)}^{(\tilde{x}, \tilde{u})} - I\| = \|L^\dagger(\hat{A}'(\tilde{u}) - \hat{A}'(u))\|$$

(c) *If $\hat{A}'(u)$ is boundedly invertible, \hat{A}' is continuous, and $\mathcal{N}(C) = \{0\}$, then for all \tilde{x} in a sufficiently small neighborhood of x , the reduced adjoint range invariance condition*

$$(55) \quad F'(\tilde{x}) = R_x^{\tilde{x}} F'(x)$$

is satisfied with

$$\|R_x^{\tilde{x}} - I\| = \sqrt{\|C(\hat{A}'(S(x)) - \hat{A}'(S(\tilde{x})))\hat{A}'(S(\tilde{x}))^{-1}C^\dagger\|^2 + \|\text{Proj}_{\mathcal{R}(C)^\perp}\|^2}$$

(d) *If $\hat{A}'(u)$ is boundedly invertible, \hat{A}' is continuous, and $\overline{\mathcal{R}(L)} = W^*$, then the reduced range invariance condition*

$$(56) \quad F'(\tilde{x}) = F'(x)R_x^{\tilde{x}}$$

is satisfied with

$$\|R_x^{\tilde{x}} - I\| = \sqrt{\|L^\dagger(\hat{A}'(S(x)) - \hat{A}'(S(\tilde{x})))\hat{A}'(S(\tilde{x}))^{-1}L\|^2 + \|\text{Proj}_{\mathcal{N}(L)}\|^2}$$

ξ	τ^2	it _{ao}	it _{red}	cpu _{ao}	cpu _{red}	$\frac{\ b_{k_*}^\delta(\delta), \text{ao} - b^\dagger\ _X}{\ b^\dagger\ _X}$	$\frac{\ b_{k_*}^\delta(\delta), \text{red} - b^\dagger\ _X}{\ b^\dagger\ _X}$
0	4	34	32	0.14	0.10	0.0149	0.0151
10	20	43	43	0.20	0.55	0.0996	0.1505
100	20	55	56	0.28	0.82	0.0721	0.0770
1000	20	68	68	0.42	1.07	0.0543	0.0588
-0.5	4	33	32	0.13	0.35	0.1174	0.2165
-1.	4	35	-	0.23	-	0.2023	-
-10	4	44	-	0.23	-	0.0768	-
-100	4	77	-	0.59	-	0.2246	-
-1000	4	70	-	0.49	-	0.0321	-

TABLE 1

Comparison of reduced and all-at-once IRGNM

Proof. Since the derivatives here simplify to $A_x(x, u) = L$, $A_u(x, u) = \hat{A}'(u)$, $C'(u) = C$, the assertions can be readily checked by setting

- (a) $R_{11} = I$, $R_{12} = (\hat{A}'(\tilde{u}) - \hat{A}'(u))C^\dagger$, $R_{21} = 0$, $R_{22} = I$
- (b) $R_{11} = I$, $R_{12} = L^\dagger(\hat{A}'(\tilde{u}) - \hat{A}'(u))$, $R_{21} = 0$, $R_{22} = I$
- (d) $R_{\tilde{x}} = C\hat{A}'(S(x))\hat{A}'(S(\tilde{x}))^{-1}C^\dagger$
- (c) $R_{\tilde{x}} = L^\dagger\hat{A}'(S(x))\hat{A}'(S(\tilde{x}))^{-1}L$

□

The above example (54) fits into this framework with

$$X = L^2(\Omega), \quad V = H^2(\Omega) \cap H_0^1(\Omega), \quad W = L^2(\Omega), \quad Y = L^2(\Omega)$$

$$\hat{A}(u) = -\Delta u + \zeta u^3, \quad L = -\text{id}$$

and therefore satisfies (43), and in the elliptic case $\xi > \xi = -\|\text{id}_{H_0^1(\Omega) \rightarrow L^4(\Omega)}\|$ also (56), provided C is linear. In case of full measurements relative to the model A , i.e., if $\mathcal{N}(C) \subseteq \mathcal{N}(\hat{A}'(u))$ for all $(x, u) \in \mathcal{D}(\mathbf{F}) \cap \mathcal{B}_\rho(x_0, u_0)$, also (41) holds and if $\xi > \xi$ and even $\mathcal{N}(C) = \{0\}$, $\overline{\mathcal{R}(C)} = Y$ hold, then (55) is satisfied.

5.2. Numerical experiments. In the following, we show numerical results for this example in the quadratic (i.e., $m = o = r = 2$) Hilbert space setting in the one-dimensional situation $\Omega = (0, 1)$ with a simple finite difference discretization on an equidistant grid of size 0.01. Computational results in 2-d with an adaptive discretization can be found in [14].

Table 1 shows a comparison of the reduced and the all-at-once versions of the IRGNM for different values of the nonlinearity parameter ξ with one per cent noise in the data. Here it_{ao}, it_{red}, cpu_{ao}, cpu_{red} denote the number of iterations and CPU times (in seconds) for the all-at-once and the reduced version, respectively. The corresponding reconstructions are displayed in Figure 1. In both cases the sequence of regularization parameters was chosen as $\alpha_k = 10 \cdot 0.7^k$ and the constant functions with value zero were used as starting values for b (and u). As expected, for larger values of ξ , i.e., stronger nonlinearity, the all-at-once version performs better. Also for negative values of ξ with modulus larger than the reciprocal norm of the embedding $H_0^1(\Omega) \rightarrow L^4(\Omega)$, for which the parameter-to-state-map does not exist (and the reduced IRGNM actually fails) we still get a reasonable behavior of the all-at-once IRGNM.

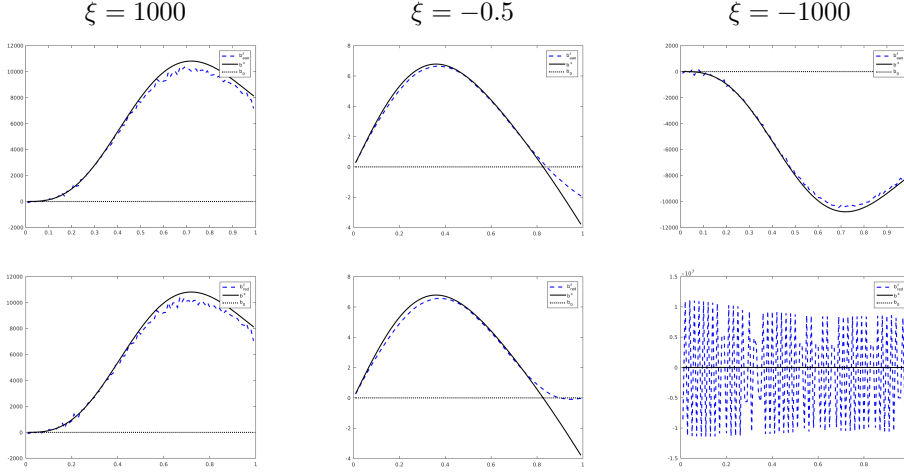


FIG. 1. Comparison of reduced (bottom row) and all-at-once (top row) IRGNM

ξ	τ^2	it _{aao}	it _{red}	cpu _{aao}	cpu _{red}	$\frac{\ b_{k_*(\delta),aao}^\delta - b^\dagger\ _X}{\ b^\dagger\ _X}$	$\frac{\ b_{k_*(\delta),red}^\delta - b^\dagger\ _X}{\ b^\dagger\ _X}$
0.5	4	5178	2697	2.97	18.07	0.0724	0.1047
5	4	2000000	48510	1293.60	482.19	0.7837	0.1633
10	4	2000000	100000	1257.50	639.87	0.9621	0.1632
-0.5	4	10895	2016	8.85	14.55	0.1406	0.2295
-1	4	18954	-	11.42	-	0.2313	-

TABLE 2
Comparison of reduced and all-at-once Landweber

A similar comparison is done for the Landweber iteration in Table 2 and Figure 2, however, with less choices of ξ since here already for a relatively moderate (positive) nonlinearity the all-at-once Landweber iteration fails, which is no surprise, since it does not contain solution of models at all. For negative values of ξ , like in the IRGNM tests, the reduced version only works with sufficiently small $|\xi|$, whereas the all-at-once version is also able to cope with absolutely slightly larger negative values of ξ .

6. Further remarks on the comparison between all-at-once and reduced version.

6.1. Comparison of conditions on forward operator. Assumptions 2, 3, and 6 cannot be directly compared with their respective reduced versions in the sense that one of them would imply the other. They have to be checked on a case by case basis, which is beyond the scope of this paper but will be subject of future research, e.g., for the examples from the introduction.

As far as the more structural conditions on nonlinearity of the forward operator (tangential cone condition as well as range invariance and adjoint range invariance) are concerned, they seem to be quite different in the reduced and in the all-at-once setting. For instance, both conditions have been proven to hold for certain special cases of Example 1 in the reduced setting see, e.g., [5, 9, 15] but they do not seem to be rigorously verifiable for these cases in the all-at-once setting. To demonstrate this for the adjoint range invariance condition (41) we consider the two special cases

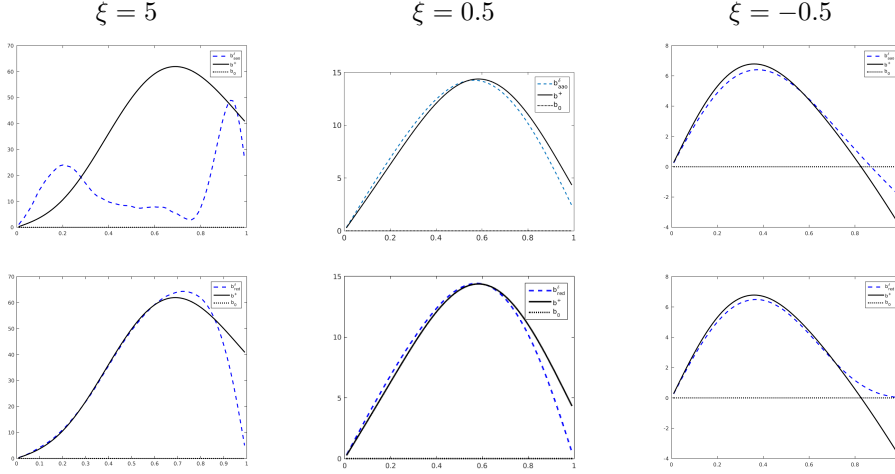


FIG. 2. Comparison of reduced (bottom row) and all-at-once (top row) Landweber

- (i) identify c , while $a \equiv 1$, $b \in H^{-1}(\Omega)$ given, (the so-called c problem), $d \in \{1, 2, 3\}$;
 - (ii) identify a , while $b \in H^{-1}(\Omega)$, $c \equiv 0$ given (the so-called a problem), $d = 1$;
- with full observations, i.e., C the embedding operator $V \rightarrow Y$. Note that these are exactly the cases in which the analogous adjoint range invariance condition can be verified for the reduced formulation.

For simplicity of exposition we switch to inhomogeneous Dirichlet instead of Neumann boundary conditions, i.e.,

$$-\nabla(a\nabla\hat{u}) + c\hat{u} = b \text{ in } \Omega, \quad \hat{u} = g_\Gamma \text{ on } \partial\Omega,$$

which with the harmonic extension $g \in H^1(\Omega)$ of the inhomogeneous Dirichlet boundary data g_Γ to the interior and $\hat{u} = u + g$, can be formulated variationally as

$$u \in H_0^1(\Omega) \text{ and for all } v \in H_0^1(\Omega) : \int_{\Omega} (a\nabla(u+g)\nabla v + c(u+g)v) = 0$$

where we assume g_Γ to be chosen such that

$$|u + g| \geq \hat{c} > 0 \text{ in } \Omega \text{ in case (i) and } |(u + g)'| \geq \hat{c} > 0 \text{ in } \Omega \text{ in case (ii)}$$

for all $(x, u) \in \mathcal{D}(\mathbf{F}) \cap \mathcal{B}_\varepsilon(x_0, u_0)$, a condition which can be incorporated into the definition of $\mathcal{D}(\mathbf{F})$.

For the c problem (i) with $x = c$ we make use of higher elliptic regularity and consider the function space setting

$$X = L^\infty(\Omega), \quad V = H^2(\Omega) \cap H_0^1(\Omega), \quad W = L^2(\Omega), \quad Y = L^p(\Omega)$$

with $p \in [1, \infty]$. We have $A_x(x, u)h = h(u + g)$, $A_u(x, u)v = -\Delta v + xv$ and set

$$\begin{aligned} R_{11}w^* &= \frac{\tilde{u} + g}{u + g}w^* & R_{12}y &= \frac{u - \tilde{u}}{u + g}(-\Delta y) + \frac{\tilde{x}(u + g) - x(\tilde{u} + g)}{u + g}y \\ R_{21} &= 0 & R_{22} &= I \end{aligned}$$

for $w^* \in L^2(\Omega)$, $y \in L^p(\Omega)$. to formally obtain (41). It remains to bound the difference $\|\mathbf{R}_{(x,u)}^{(\tilde{x},\tilde{u})} - I\| = \sqrt{\|R_{11} - I\|_{W^* \rightarrow W^*}^2 + \|R_{12}\|_{Y \rightarrow W^*}^2}$, which we do by estimating

$$\begin{aligned} \|R_{11}w^* - w^*\|_{W^*} &= \left\| \frac{\tilde{u} - u}{u + g} w^* \right\|_{L^2(\Omega)} \leq \left\| \frac{\tilde{u} - u}{u + g} \right\|_{L^\infty(\Omega)} \|w^*\|_{L^2(\Omega)} \\ &\leq \frac{C_{H^2 \rightarrow L^\infty}}{\hat{c}} \|\tilde{u} - u\|_V \|w^*\|_{W^*} \end{aligned}$$

for any $w^* \in L^2(\Omega)$ and

$$\|R_{12}y\|_{W^*} = \left\| \frac{u - \tilde{u}}{u + g} (-\Delta y) + \frac{\tilde{x}(u + g) - x(\tilde{u} + g)}{u + g} y \right\|_{L^2(\Omega)}$$

for any $y \in L^p(\Omega)$, which we are supposed to estimate by a small multiple of $\|y\|_{L^p(\Omega)}$. However, the appearance of $-\Delta y$ under the $L^2(\Omega)$ norm indicates that although the identity (41) formally holds, the required estimate on $\|R - I\|$ in (40) does not seem to be obtainable.

For the 1-d a problem (ii) with the somewhat more convenient boundary conditions

$$-(a\hat{u}')' = 0 \text{ on } (0, 1), \quad \hat{u}'(0) = g^1, \quad \hat{u}(1) = g_1,$$

and under the assumption that a takes the known value a_0 at the left hand boundary point, we use the extension $g(s) = g^1(s - 1) + g_1$, $s \in (0, 1)$ and the Ansatz $a = a_0 + x$, $\hat{u} = u + g$. With

$$\begin{aligned} X &= \{x \in W^{1,q}(0, 1) : x(0) = 0\}, \quad V = \{x \in W^{2,q}(0, 1) \cap H^1(0, 1) : v'(0) = 0\}, \\ W^* &= L^q(0, 1), \quad Y = L^p(0, 1), \end{aligned}$$

where the prescribed boundary values are to be understood in a trace sense and $p \in [1, \infty]$, $q \in (1, \infty)$, $A(x, u) = -(a_0 + x)u'' - x'(g^1 + u')$, $A_x(x, u)h = -hu'' - h'(g^1 + u')$, $A_u(x, u)v = -(a_0 + x)v'' - x'v'$ and setting

$$\begin{aligned} R_{11}w^* &= \left(\frac{(\tilde{u} + g)'}{(u + g)'} \int_0^\cdot w^* ds \right)' & R_{12}y &= - \left(\frac{(a_0 + \tilde{x})(u + g)' - (a_0 + x)(\tilde{u} + g)'}{(u + g)'} y' \right)' \\ R_{21} &= 0 & R_{22} &= I \end{aligned}$$

we formally satisfy (41). Also here the estimate of the $R_{11} - I$ term of $\mathbf{R} - I$ works out

$$\begin{aligned} \|R_{11}w^* - w^*\|_{W^*} &= \left\| \left(\frac{\tilde{u}''(g^1 + u') - u''(g^1 + \tilde{u}')}{(g^1 + u')^2} \int_0^\cdot w^* ds \right)' + \frac{\tilde{u}' - u'}{g^1 + u'} w^* \right\|_{L^q(0,1)} \\ &\leq \left(\frac{1}{\hat{c}} \|\tilde{u}'' - u''\|_{L^q(0,1)} + \frac{1}{\hat{c}^2} \|u''\|_{L^q(0,1)} \|\tilde{u}' - u'\|_{L^\infty(0,1)} \right) \|w^*\|_{L^1(0,1)} \\ &\quad + \frac{1}{\hat{c}} \|\tilde{u}' - u'\|_{L^\infty(0,1)} \|w^*\|_{L^q(0,1)}, \end{aligned}$$

which by embeddings can be estimated by some constant times $\|\tilde{u} - u\|_V \|w^*\|_{W^*}$, whereas $R_{12}y$ contains derivatives of y that prevent an estimate of its $L^q(0, 1)$ norm by a small multiple of $\|y\|_{L^p(0,1)}$.

6.2. Comparison of source conditions in reduced and all-at-once setting. As a consequence of Proposition (3) we expect the source condition (22) to be equivalent for the reduced and the all-at-once version of Tikhonov if $m = 1$ (and ρ is large enough). However, for $m > 1$, at a first glance there might be a difference, so we now consider the special case of $m = o = r = 2$ in a Hilbert space setting and compare the reduced and the all-at-once benchmark source condition, i.e., the one yielding $O(\sqrt{\delta})$ convergence of the error norm in both versions. The reduced source condition

$$x^\dagger - x_0 = R_X F'(x^\dagger)^* v$$

for some $v \in Y^*$ with the abbreviations $L = A_x(x^\dagger, u^\dagger)$, $K = A_u(x^\dagger, u^\dagger)$ and the Riesz isomorphism $R_X : X^* \rightarrow X$ is equivalent to

$$(57) \quad x^\dagger - x_0 = -R_X(K^{-1}L)^* C'(u^\dagger)^* v$$

whereas the all-at-once version (22)

$$(58) \quad \begin{pmatrix} x^\dagger - x_0 \\ u^\dagger - u_0 \end{pmatrix} = R_{X \times V} \mathbf{F}'(x^\dagger, u^\dagger)^* \begin{pmatrix} v^{\text{mod}} \\ v^{\text{obs}} \end{pmatrix} = \begin{pmatrix} R_X L^* v^{\text{mod}} \\ R_V(K^* v^{\text{mod}} + C'(u^\dagger)^* v^{\text{obs}}) \end{pmatrix}$$

for some $(v^{\text{mod}}, v^{\text{obs}}) \in W \times Y^*$ after elimination of v^{mod} (relying on Assumption 1) and using $u^\dagger = S(x^\dagger)$, $-K^{-1}L = S'(x^\dagger)$ can be rewritten as

$$(59) \quad \begin{aligned} v^{\text{mod}} &= (K^*)^{-1} \left(R_V^{-1}(u^\dagger - u_0) - C'(u^\dagger)^* v^{\text{obs}} \right) \\ x^\dagger - x_0 + R_X((K^{-1}L)^* R_V^{-1}(K^{-1}Lx^\dagger + u_0)) &= -R_X(K^{-1}L)^* C'(u^\dagger)^* v^{\text{obs}}. \end{aligned}$$

In the linear case $A_x \equiv L$, $A_u \equiv K$, with $S = -K^{-1}L$, and the positive definite operator $T = I + R_X S^* R_V^{-1} S$ the latter is equivalent to

$$(60) \quad T(x^\dagger - x_0) = -R_X(K^{-1}L)^* (C'(u^\dagger)^* v^{\text{obs}} + R_V^{-1}(K^{-1}Lx_0 + u_0)).$$

Thus, replacing the linear model $Ku + Lx = 0$ by its transformed version $Ku + LT^{-1}\tilde{x} = 0$, we see that both source conditions (57) and (60) are equivalent in the Hilbert space case (even with possibly nonlinear observations) provided the initial point satisfies the model. This is actually not surprising in view of well-known converse results for linear inverse problems cf., [6, Section 4.2] and the fact that, e.g., the respective version of Tikhonov regularization yields the same $O(\sqrt{\delta})$ convergence rate under both conditions.

However, there might still be a considerable difference in the nonlinear and/or Banach space case setting with variational source conditions. This conjecture is supported by the observation that the nonconvergence at the right endpoint of the interval for the respective reduced version in case $\xi = -0.5$ of Figure 1 and cases $\xi = \pm 0.5$ of Figure 2 (as expected since the difference between x^\dagger and x_0 in this point is known to lead to violation of the reduced benchmark source condition) seems to be relaxed in the all-at-once versions.

7. Conclusions and Outlook. All-at-once versions of regularization methods can offer advantages over their classical reduced counterparts when it comes to avoiding explicit use of parameter-to-state maps, i.e., of exactly solving possibly nonlinear models in each step of iterative methods. More precisely, while there is no significant difference in the implementation of all-at-once and reduced Tikhonov regularization,

in reduced Newton iterations one has to solve nonlinear and linear models in each step, while an all-at-once Newton step only requires to solve linearized models. Still going further, as opposed to reduced Landweber, which amounts to solving a nonlinear and an adjoint linear model in each step, there is no model solved at all in an all-at-once Landweber step.

It remains to more thoroughly compare source conditions and restrictions on the nonlinearity like tangential cone and (adjoint) range invariance conditions for some relevant model problems and for real applications. Moreover we will investigate more general data misfit and regularization functionals, as well as other regularization paradigms.

REFERENCES

- [1] A. BAKUSHINSKII, *The problem of the convergence of the iteratively regularized Gauß-Newton method*, Comput. Maths. Math. Phys., 32 (1992), pp. 1353–1359.
- [2] A. BAKUSHINSKY AND M. KOKURIN, *Iterative Methods for Approximate Solution of Inverse Problems*, vol. 577 of Mathematics and Its Applications, Springer, Dordrecht, 2004.
- [3] M. BURGER AND W. MÜHLHUBER, *Iterative regularization of parameter identification problems by sequential quadratic programming methods*, Inverse Problems, 18 (2002), pp. 943–969.
- [4] M. BURGER AND W. MÜHLHUBER, *Numerical approximation of an SQP-type method for parameter identification*, SIAM J. Numer. Anal., 40 (2002), pp. 1775–1797.
- [5] H. ENGL, K. KUNISCH, AND A. NEUBAUER, *Convergence rates for Tikhonov regularisation of non-linear ill-posed problems*, Inverse Problems, 5 (1989), pp. 523–540.
- [6] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.
- [7] J. FLEMMING, *Generalized Tikhonov Regularization: Basic Theory and Comprehensive Results on Convergence Rates*, Shaker Verlag, Aachen, 2012. PhD thesis, TU Chemnitz 2011, <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-78152>.
- [8] E. HABER AND U. M. ASCHER, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Inverse Problems, 17 (2001), p. 1847.
- [9] M. HANKE, A. NEUBAUER, AND O. SCHERZER, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), pp. 21–37.
- [10] B. HOFMANN, B. KALTENBACHER, C. PÖSCHL, AND O. SCHERZER, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.
- [11] T. HOHAGE AND F. WERNER, *Iteratively regularized Newton methods with general data misfit functionals and applications to Poisson data*, Numerische Mathematik, 123 (2013), pp. 745–779.
- [12] B. KALTENBACHER, *Some Newton type methods for the regularization of nonlinear ill-posed problems*, Inverse Problems, 13 (1997), pp. 729–753.
- [13] B. KALTENBACHER AND B. HOFMANN, *Convergence rates for the iteratively regularized Gauss-Newton method in Banach spaces*, Inverse Problems, 26 (2010), p. 035007.
- [14] B. KALTENBACHER, A. KIRCHNER, AND B. VEXLER, *Goal oriented adaptivity in the IRGNM for parameter identification in PDEs II: all-at once formulations*, Inverse Problems, (2014), p. 045002.
- [15] B. KALTENBACHER, A. NEUBAUER, AND O. SCHERZER, *Iterative regularization methods for nonlinear ill-posed problems*, vol. 6 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter GmbH & Co. KG, Berlin, 2008.
- [16] B. KALTENBACHER, F. SCHÖPFER, AND T. SCHUSTER, *Convergence of some iterative methods for the regularization of nonlinear ill-posed problems in Banach spaces*, Inverse Problems, 25 (2009). 065003 doi: 10.1088/0266-5611/25/6/065003.
- [17] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, New York, 1996.
- [18] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering, Springer New York, 2006, <https://books.google.at/books?id=VbHYoSyeIFcC>.
- [19] C. PÖSCHL, *Tikhonov regularization with general residual term*, PhD thesis, University of Innsbruck, 2008.
- [20] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, AND F. LENZEN, *Variational*

- Methods in Imaging*, Springer, New York, 2008.
- [21] T. SCHUSTER, B. KALTENBACHER, B. HOFMANN, AND K. KAZIMIERSKI, *Regularization Methods in Banach Spaces*, de Gruyter, Berlin, New York, 2012. Radon Series on Computational and Applied Mathematics.
 - [22] F. TRÖLTZSCH, *Optimal control of partial differential equations*, vol. 112, American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.
 - [23] T. VAN LEEUWEN AND F. J. HERRMANN, *A penalty method for pde-constrained optimization in inverse problems*, Inverse Problems, 32 (2016), p. 015007, <http://stacks.iop.org/0266-5611/32/i=1/a=015007>.
 - [24] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Providence, 2002.
 - [25] F. WERNER, *Inverse problems with Poisson data: Tikhonov-type regularization and iteratively regularized Newton methods*, Der andere Verlag, Uelvesbüll, 2012. PhD thesis, University of Göttingen 2012, http://num.math.uni-goettingen.de/~f.werner/files/diss.frank_werner.pdf.
 - [26] F. WERNER, *On convergence rates for iteratively regularized newton-type methods under a lipschitz-type nonlinearity condition*, Journal of Inverse and Ill-Posed Problems, (2015).